Construction of a genomic DNA 'feature map' by sequencing from nested deletions: application to the HLA class I region

B. Rajendra Krishnan¹, Isabela Jamry², Douglas E. Berg³, Claire M. Berg⁴ and David D. Chaplin^{1-3,*}

¹Division of Allergy and Immunology, Department of Internal Medicine, ²Howard Hughes Medical Institute and ³Department of Molecular Microbiology, Washington University School of Medicine, St Louis, MO 63110, USA and ⁴Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT 06269, USA

Received September 22, 1994; Revised and Accepted November 28, 1994

GenBank accession nos U18138-U18195 (incl.)

ABSTRACT

We are applying a transposon-based approach for detecting and mapping features of special interest to construct 'feature maps' of currently uncharacterized portions of the human leukocyte antigen (HLA) complex on chromosome 6. Such feature maps should facilitate identifying regions for high resolution analysis. Here we describe the feature mapping of a 35 kb DNA fragment located between the *HLA-C* and *HLA-E* loci. This fragment was cloned into a transposon $\gamma\delta$ -based cosmid vector designed for generating nested deletions *in vivo*. Seventy informative nested deletions extending into the cloned fragment were isolated, and DNA adjacent to the deletion endpoints was sequenced by fluorescent automated technology. These islands of DNA sequences constituted the foundation of the feature map, and (i) identified putative exons, (ii) determined the positions of *Alu* elements, (iii) determined the span of the keratinocyte-specific *S* gene, and (iv) localized evolutionarily conserved sequences. The construction of feature maps using this *in vivo* nested deletion-sequencing approach provides a rapid and efficient means to identify DNA regions that merit more detailed analysis.

INTRODUCTION

It is often desirable in large scale genome analysis to locate and map features of special interest ('feature mapping') to help guide the choice of regions to be subjected to detailed analyses. Feature maps, constructed from islands of sequences at known locations, contain the positions of known or potential genes, repetitive sequences, regions with homology to sequences in the database, and evolutionarily conserved sequences. They can serve to focus attention on subregions particularly likely to contain previously unrecognized genes. Strategies incorporating feature mapping should increase the efficiency of analysis of large genomic

regions. Effective use of this strategy depends on knowledge of positions of the sequence islands within the large DNA regions under study.

The human leukocyte antigen (HLA) complex spans ~4 megabases (Mb) on the short arm of chromosome 6. Portions of the HLA Class I region are gene dense, with a gene every ~20 kb of genomic DNA (1,2). However, in an ~160 kb DNA region near the *HLA-C* locus, only one gene (the keratinocyte-specific *S* gene) has been identified (3).

We have developed transposon $\gamma\delta$ -based cosmid cloning vectors for generating nested deletions that extend into large cloned DNA segments by *in vivo* transposition and simple bacteriological selection ('deletion factory vectors'; 4,5; Fig. 1). These deletions place primer binding sites throughout the DNA of interest at locations that are easily determined by plasmid size. Here we have applied this approach to a 35 kb DNA region ~150 kb telomeric of *HLA-C*. The DNA sequence adjacent to each of 70 deletion endpoints was determined and data from these sequence islands were used to construct a feature map of this 35 kb DNA.

MATERIALS AND METHODS

Bacterial strains, plasmids and bacteriological techniques

E. coli DH10B (6) was used as the standard strain for construction of recombinant plasmids and for preparing plasmid DNA for sequencing. To isolate nested deletions, DH10B containing the 35 kb fragment cloned in the 'deletion factory vector', pDELTA1 (4,5), was transformed with plasmid pXRD4043, which encodes the $\gamma\delta$ -transposase (7).

Luria broth (8) was used as the standard bacterial growth medium. Solid media contained 1.5% (w/v) agar (Difco). Concentrations of selection agents in solid media were: ampicillin (Ap), 100 μ g/ml; chloramphenicol (Cm), 20 μ g/ml; kanamycin (Km), 50 μ g/ml; streptomycin (Sm), 100 μ g/ml; sucrose (Suc), 5% (w/v); and tetracycline (Tc), 20 μ g/ml. One-half of these concentrations were used in liquid media.

* To whom correspondence should be addressed at: Division of Allergy and Immunology, Department of Internal Medicine, Washington University School of Medicine, Box 8122, 660 South Euclid Avenue, St Louis, MO 63110, USA.

Figure 1. Schematic representation of hypothetical nested deletions obtained using cosmid pDELTA1 and their phenotypes (adapted from reference 32). (A) Depiction of pDELTA1 hypothetical DNA clone. (B) Deletions extending into various sites within the cloned fragment from SucTc selection. (C) Deletions extending into various sites within the cloned fragment from SmKm selection. The shaded portion represents the cloned DNA, the filled arrowheads represent transposon $\gamma\delta$ sequences, and the half-arrows represent locations of sequencing primer binding sites (see text and references 4,5). R/r indicates resistance and S/s indicates sensitivity.

Standard methods for recombinant DNA construction, restriction analysis and hybridization were employed (8,9). Restriction enzymes and T4 DNA ligase were from New England Biolabs (Beverly, MA), Taq DyeDeoxy Terminator kit from Applied Biosystems Inc. (ABI, Foster City, CA), and Centri-Sep columns from Princeton Separations (Adelphi, NJ). Other reagents were from Life Technologies Inc. (Gaithersburg, MD), or International Biotechnologies Inc. (New Haven, CT).

Electrocompetent *E. coli* were prepared by growing cells with vigorous shaking at 37°C in 250 ml of 2× YT medium (0.5% w/v sodium chloride, 1% w/v yeast extract and 2% w/v Bacto-tryptone) to OD₆₀₀ = 0.6, washing three times in sterile ice-cold 10% glycerol and concentrating 100-fold in 10% glycerol. Competent cells were stored at -70°C and 50 µl aliquots were used for electroporation. Electroporation was done using a Bio-rad Gene-Pulser apparatus at 1.8 kV, 200 ohms and 25 µFD. The cells were rescued in SOC medium (9) and plated on appropriate selection plates after incubation at 37°C for 60 min.

Cloning into pDELTA1 and selection of deletions

A 35 kb DNA fragment containing HLA sequences was recovered from a cosmid clone, L14 (containing a subfragment of YAC, yB209D7; reference 10), by digestion with *NorI* and ligated with *NorI*-digested and dephosphorylated pDELTA1 DNA (5). The heat-treated (65°C, 15 min) ligation mixture was used to electroporate *E. coli* DH10B and transformants were selected on medium containing Km, Tc and X-gal (5-bromo-4-chloro-3-indolyl-β-D galactoside; 2% w/v in dime-

thylformamide). Several white colonies were restreaked on the same medium and analyzed for plasmid content. Two plasmids containing the 35 kb fragment cloned in pDELTA1 in opposite orientations (pL14DEL1 and pL14DEL2, respectively) were chosen for further analyses.

General protocols for selecting deletions using pDELTA1 and related pDUAL deletion factory vectors have been described (4,5). Briefly, five colonies carrying the appropriate cloned DNA were pooled and inoculated in 5 ml of Luria broth containing Cm, Km and Tc and incubated at 37°C with aeration for 6 h. Then, 0.04 ml was inoculated into 2 ml of Luria Broth containing just Cm and Km or just Cm and Tc, and incubated for a further 6 h with shaking at 37°C. Plasmid DNA was purified immediately from these cultures and electroporated into *E. coli* DH10B. Transformants from Cm plus Km cultures were selected for resistance to Sm (loss of *strA*) and Km, and transformants from Cm plus Tc cultures were selected for resistance to sucrose (loss of *sacB*) and Tc (see Fig. 1). Deletions in these vectors are formed by $\gamma\delta$ -mediated intramolecular transposition, and a molecular model for deletion formation has been described elsewhere (4).

Plasmid DNA preparation, analysis and fluorescent sequencing

To obtain initial estimates of plasmid size, supercoiled plasmid DNA was electrophoresed. Bacterial colonies were resuspended in 25 µl of TE in microtiter plates using sterile toothpicks, and 25 µl of lysis solution (10% w/v sucrose, 0.2 N NaOH, 1% w/v SDS) was added. After 3 min at room temperature, 5 µl of 4 M KCl was

added, the plates were chilled on ice, and 5 μ l of Stop dye (95% formamide, 20 mM EDTA, 0.05% w/v bromophenol blue, 0.05% w/v xylene cyanol) was added to each sample. After 5 min on ice, the microtiter plates were centrifuged at 1200 r.p.m. for 5 min in a Beckman J6-M centrifuge and 40 μ l of supernatants were electrophoresed in 0.7% agarose gels. More precise determinations of plasmid sizes, and thus deletion endpoint positions, were obtained by digestion with *NotI* to cleave at one vector fragment junction, or by digestion with *Bam*HI, *Eco*RI, *Hind*III or *Pst*I and electrophoresis in 0.7% agarose gels.

To isolate plasmid DNA for sequencing, the Qiawell-8 DNA preparation system (Qiagen Inc., CA) was used. Cycle sequencing was performed using labeled dideoxynTPs and the Taq DyeDeoxy Terminator kit from ABI in a Perkin-Elmer TC480 thermal cycler. To obtain reliable sequence information we found it necessary to use longer primers than those recommended previously (5,11): the primer used to sequence the SmKm deletion derivatives was T7lp1 (5' GCGGCCGAGCTCACACAGAT) and the primer used to sequence the SucTc deletion derivatives was SP6lp (5' GCGTCCGATTTAGGTGACACTATAGA). Sequencing reaction products were resolved on an ABI 370A sequencer updated with the 373A 1.2.0 software (11).

Computer-assisted sequence analysis

Overlap detection and 'contig' building were done using the Geneworks software (version 2.2) from Intelligenetics Inc. Screening for potential exons was done by 'GRAIL' analysis (12) and homology searches to the databases were performed by 'BLAST' analysis (13,14). Sequences that did not exhibit homology to the database by BLAST analysis were also examined by 'FASTA' analysis (15). HLA sequences reported in this paper have been submitted to GenBank, and assigned accession numbers U18138–U18195 (incl.).

RESULTS

Nested deletion derivatives of pL14DEL1 and pL14DEL2

Nested deletions were generated by $\gamma\delta$ transposition *in vivo* in plasmid pL14DEL1, which contains a 35 kb DNA fragment cloned in pDELTA1. For a first pass screening, and to assess whether deletion endpoints are distributed randomly, we analyzed 50 colonies plated on SmKm and 50 colonies plated on SucTc. Sizes of the plasmids were estimated by electrophoresing supercoiled DNA directly, and also following *NotI* digestion. The sizes of deletion derivatives ranged from 5.1 to 23 kb (Fig. 2). From these, the following 26 deletions were chosen for further analysis: A4, A6–A8, A10, A12–A16 and A21 (derived by SmKm selection) and B4, B5, B8, B11, B12, B15, B16, B18–B20, B22–B24, B26 and B28 (derived by SucTc selection) (marked by asterisks in Fig. 3). Deletions designated A contain the left (centromere-proximal) end of the HLA-specific DNA and those designated B contain the right (centromere-distal) end of the HLA-specific DNA (see Fig. 3).

To fill gaps remaining in the first pass screening, we next selected deletions in pL14DEL2, which contains the same 35 kb HLA DNA cloned in the opposite orientation (relative to pL14DEL1) in pDELTA1. Seventy five random colonies from the SmKm selection and 75 random colonies from the SucTc selection were screened and the following deletion derivatives

Figure 2. *NotI* linearized DNA of the B series. Lanes 2–29 contained *NotI*-digested DNA from plasmids of the B series (Fig. 3): lanes 2–19 contained DNA from plasmids B1–B18; lane 20 contained DNA from plasmid B23; lane 21 contained DNA from plasmid B20; lanes 22 and 23 contained DNA from plasmids B26 and B27; lanes 24 and 25 contained DNA from plasmids B24 and B25; lanes 26 and 27 contained DNA from plasmids B28 and B29; lane 28 contained DNA from plasmid B32; and lane 29 contained DNA from plasmid B30, respectively. Lanes 1 and 30 contained DNA molecular weight markers (1 kb DNA ladder and *Hind*III-digested phage λ DNA, respectively; sizes (kb) of selected bands are shown on the sides of the gel).

were mapped: A1, A2, A19, A26 and A28–A38 (derived by SucTc selection) and B1–B3, B6, B7, B9, B10, B13, 14, B17, B21, B25, B27 and B29–32 (derived by SmKm selection). Further screening of pL14DEL1 derivatives yielded A3, A5, A9, A11, A17, 18, A20 and A22–25 (derived by SmKm selection). Cumulatively, these 70 derivatives constituted a nested series with a uniform distribution of deletion endpoints (Fig. 3). Accurate sizing and ordering of the nested deletion derivatives was facilitated by restriction enzyme digestion using endonucleases *Bam*HI, *Eco*RI, *Hind*III, and *Pst*I (data not shown). Each was then used for sequencing from the appropriate transposon end: 250–325 bp of reliable sequence were routinely obtained, independent of plasmid size in the 5–23 kb size range (11).

'Feature mapping' of HLA DNA using nested deletions

Assembly of a deletion series and identification of features. We expected that sequence signatures of a few hundred bp located every ~2 kb would be sufficient to find most important features in a DNA segment. Consequently, we sought ~20 uniformly spaced nested deletions. Size analysis of the 26 deletion derivatives assembled during the first pass screening indicated no overlap between deletions from opposite orientations, and also a lack of deletions at the left end of the DNA in L14 (Fig. 3). Accordingly, additional deletions from both pL14DEL1 and pL14DEL2 were mapped: of these, 44 were used to identify deletion derivatives spanning the left end of the DNA in L14, to yield an overlapping set of nested deletions and to refine the feature map of the 35 kb DNA fragment (see Fig. 3). When the 70 sequence fragments were analyzed for 'features' (putative exons, *Alu* elements, and matches to the database), 7 were found to contain putative exons (Table 1), 12 to contain homology to *Alu* sequences and 1 to exhibit homology to an entry in the database (see below).

Figure 3. Feature map of L14 HLA DNA and summary of deletion derivatives obtained in this study. The relationship of L14 HLA DNA with respect to its parental YAC, yB209D7 and to the Class I region on chromosome 6 are shown at the top. Numbers on the left (A1–A38) and right (B1–B32) indicate the deletion plasmids. Asterisks indicate plasmids obtained in the first pass screening (see text for details). Features of the 35 kb L14 HLA DNA are indicated in the figure as follows: inverted Ts indicate locations of putative exons (by GRAIL analysis; see also Table 1); filled rightward pointing arrow indicates the *S* gene and direction of its open reading frame; shaded boxes indicate the evolutionarily conserved 0.6 and 3.4 kb DNA regions; hatched boxes indicate the locations of *Alu* elements; and the cross-hatched box indicates YAC vector-derived sequences.

Prediction of exons. Sequences derived from 7 of the deletion plasmids scored as 'good' or 'excellent' concerning exon character by GRAIL analysis (Table 1 and Fig. 3). One of these (sequence obtained from B5) maps to the 5' end of the *S* gene (see below). Interestingly, GRAIL analysis also predicted a cluster of five exons at the left end of the 35 kb fragment near the *Alu* repetitive elements, suggesting a currently unrecognized gene(s) in this location.

Table 1 Quality of open reading frame in sequences obtained from seven deletion plasmids by GRAIL analysis

No. ^a	Plasmid	Length of sequenced fragment (nt)	Length of predicted ORF	Likelihood of being an exon
1	A38	316	204	Good
2	A37	284	252	Good
3	A35	327	207	Excellent
4	A31	283	246	Good
5	A24	285	285	Excellent
6	B5	346	177	Good
7	B26	308	138	Excellent

^aCorresponds with the numbering of inverted Ts in Figure 3.

Defining the boundaries of the *S* gene. The sequence obtained from B11 matched the 3' untranslated portion of the *S* gene cDNA (nt positions 2166–2416; reference 3), and thereby identified the 3' end of the *S* gene (see Fig. 3). An oligonucleotide specific to exon 1 of the *S* gene (nt positions, 63–85; 5' ATGTTGG-CACTGCTGGCTGG) hybridized with deletion clones B1–B4, but not with B5–B32 (colony hybridizations performed as described in reference 8). This indicated that the 5' end of the *S* gene is also in cosmid L14, and placed it between the endpoints of the B4 and B5 deletions. Based on the positions of B4 and B5 and of the B11 deletion endpoints, we infer that the *S* gene spans ~6 kb and lies in a centromere→telomere orientation (Fig. 3). Sequences from regions immediately to the left (upstream) of the *S* gene did not exhibit significant homology to current entries in the database. Based on their physical position, we suggest that these sequences constitute *S* gene upstream regulatory elements.

Localization of *Alu* repeats. Sequences derived from A5, A6, A15, A17, A18, A21–A23, A26–A28 and A34 exhibited significant homology to *Alu* elements. These results indicate that the left half of the L14 HLA DNA contains a high density of *Alu* elements (Fig. 3). Sequences obtained from clones B22–B32 were identical to the YAC vector DNA, and located the boundary of the human DNA insert in the ancestral YAC clone, yB209D7, to the interval between B21 and B22.

Mapping evolutionarily conserved DNA regions. Two non-overlapping 0.6 and 3.4 kb *EcoRI* fragments from the L14 HLA DNA region are evolutionarily conserved (as determined by hybridization to genomic DNAs of rabbit, guinea pig, mouse and sheep; data not shown). Sequences derived from A12 contained homology to one end of the 0.6 kb region and sequences derived from B12 and B17 contained homology to the ends of the 3.4 kb region. Thus, the 0.6 kb region is ~5 kb to the left of the *S* gene,

and the 3.4 kb region is ~2 kb to the right of the *S* gene, respectively.

DISCUSSION

The HLA complex is among the best studied gene clusters in the human genome, yet large portions of it remain incompletely characterized. Among these is a 0.5 Mb region near *HLA-C* (see references 16,17). Here, we applied the nested deletion approach to build a feature map of a 35 kb subregion located ~150 kb telomeric of *HLA-C*. We (i) mapped the genomic limits of the *S* gene (a gene expressed in terminally differentiated keratinocytes), (ii) located 5 additional putative exons, (iii) mapped two evolutionarily conserved DNA fragments, and (iv) located *Alu* elements within this region. The array of deletions we generated should also be useful in further dissection of this chromosomal fragment, e.g. analysis of sequences regulating *S* gene expression.

Sequences upstream of the *S* gene from the middle portion of the cloned 35 kb fragment showed no extended homology to the database, and may constitute *S* gene regulatory elements. The left portion of the cosmid L14 insert, as depicted in Fig. 3, contains a cluster of *Alu* elements (18) and is reminiscent of a 90 kb fragment near *HLA-B* which is *Alu*-dense with clusters of *Alu* repeats flanking coding regions (19). Complete sequencing of this region of the HLA complex will fully reveal the organization and density of *Alu* repeats and their possible relationship to novel coding region(s).

For L14, we independently identified evolutionarily conserved sequences in 0.6 and 3.4 kb *EcoRI* fragments (unpublished). The nested deletion series isolated here facilitated the physical mapping of these conserved sequences to positions upstream and downstream of the *S* gene, respectively. This illustrates that ancillary data can be rapidly integrated into the feature map.

Current approaches for detailed characterization of large fragments of chromosomes remain costly and labor-intensive. It is, therefore, often more efficient to identify subregions within a chromosomal region that most strongly merit detailed analysis, e.g. by cDNA selection, exon trapping, or sequencing (20–25). Here we propose that initial characterization involve the construction of a 'feature map' consisting of islands of DNA sequence spaced every few kb along the length of the fragment. Features within the sequenced DNA may include those likely to be exons (identified by GRAIL or similar analysis) and those with homology to known sequences in the database (e.g. identified by BLAST and FASTA analyses). A major impediment to this approach is the generation of sequencing templates from large DNAs without disrupting linkage. We demonstrate here that the nested deletion approach is an efficient tool to overcome this difficulty.

Chen *et al.* (26) proposed subcloning YACs as overlapping small plasmids (~5 kb inserts), identifying subclones that overlap every 300–1000 bp, and sequencing ends of the cloned DNA in these plasmids, as a means of selecting regions of interest for 'ordered shotgun sequencing'. Their proposal requires generating high density overlap clone banks, a procedure which can be time- and labor-intensive, particularly to produce adequate numbers of uniformly spaced overlaps. Smith *et al.* (27) suggested use of a high density overlap cosmid clone bank and sequencing ends of these cloned cosmid DNAs to facilitate ordered genome analysis. In contrast to these two approaches, construction of a minimally overlapping set of cosmid subclones from a YAC requires far

fewer clones to form a minimal contig for constructing a feature map via $\gamma\delta$ -mediated nested deletions.

We hypothesize that a sequence signature every 1–2 kb of genomic DNA would be sufficient to identify the majority of features of special interest in a given genomic fragment. However, the ability to locate genes by feature mapping would depend on the gene density and gene size in the genomic fragment studied. For example, the telomeric portion of the HLA Class I region is thought to contain a gene every 20 kb of genomic DNA (1,2). Assuming that the same organization is retained in the centromeric portion (i.e. near the *HLA-C* gene), the 35 kb fragment would be predicted to harbor 2 genes. Localization of the *S* gene and mapping of an independent cluster of putative exons to the 35 kb fragment (3, this study) are consistent with this prediction.

In this study, 26 deletions (plasmids marked with asterisks in Fig. 3) obtained in the first pass screening helped identify many features in the 35 kb fragment (see Results). Analysis of an additional 44 deletions yielded uniform coverage of the entire 35 kb. The ease of selecting and ordering deletions, and the ability to sequence reliably from plasmids that vary substantially in size (11) provide a rapid and efficient means to generate feature maps of large DNAs. Furthermore, sequencing improvements enabling longer reads (28) and employing gap filling strategies (for example, via primer walking, vectorette PCR, or single site amplification; 29–31) should permit complete sequencing of most subregions using such arrays of deletions.

This study demonstrates that the transposon $\gamma\delta$ -based generation of nested deletions can be effectively applied for a sequencing-based approach to construct feature maps of cosmid-sized DNAs. Here, we exploited this method to define the location, size and organization of the *S* gene, the locations of two evolutionarily conserved DNA regions and to detect and map other exon(s) and *Alu* elements in a 35 kb DNA region near *HLA-C*. The collection of deletion derivatives generated here will also be useful in long-range structural and biological characterization of gene(s) and sequences contained in this cosmid.

ACKNOWLEDGEMENTS

We thank R. W. Blakesley for the gift of pDELTA1 and helpful discussions, and D. Schlessinger for comments on the manuscript. B. R. K. is an Arthritis Foundation Post-doctoral Fellow and D. D. C. is an investigator of the Howard Hughes Medical Institute. This study was supported by NIH grants AI15322 and AI34580 (D. D. C.), HG00563 and DK48029 (D. E. B.), and HG00820 and by the University of Connecticut Research Foundation (C. M. B.).

REFERENCES

- 1 Fan, W.-F., Wei, H., Shukla, H., Parimoo, S., Xu, H., Sankhavaram, P., Li Z., and Weissman, S.M. (1993) *Genomics* **17**, 575–581.
- 2 Wei, H., Fan, W.-F., Xu, H., Parimoo, S., Shukla, H., Chaplin, D.D., and Weissman, S.M. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 11870–11874.
- 3 Zhou, Y and Chaplin, D.D. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 9470–9474.
- 4 Wang, G., Blakesley, R.W., Berg, D.E., and Berg, C.M. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 7874–7878.
- 5 Wang, G., Berg, C.M., Chen, J., Young, A.C., Blakesley, R.W., Lee, L.Y., and Berg, D.E. (1993) *Focus* **15**, 47–49.
- 6 Grant, S.G.N., Jessee, J., Bloom, F.R., and Hanahan, D. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 4645–4649.
- 7 Tsai, M.-M., Wong, R.Y.-P., Hough, A.T., and Deonier, R.C. (1987) *J. Bacteriol.* **169**, 5556–5562.
- 8 Sambrook, J., Maniatis, T., and Fritsch, E.F. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- 9 Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K. (eds.) (1987) *Current Protocols in Molecular Biology*. Greene Publishing Associates/Wiley Interscience, New York, NY.
- 10 Bronson, S.K., Pei, J., Taillon-Miller, P., Chorney, M.J., Geraghty, D.E., and Chaplin, D.D. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 1676–1680.
- 11 Krishnan, B.R. and Chaplin, D.D. (1994) *BioTechniques* **17**, 854–857.
- 12 Uberbacher, E.C. and Mural, R.J. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 11261–11265.
- 13 Altshul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) *J. Mol. Biol.* **215**, 403–410.
- 14 Gish, W and States, D.J. (1993) *Nature Genet.* **3**, 266–272.
- 15 Pearson, W.R and Lipman, D.J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2444–2448.
- 16 Campbell, R.D. and Trowsdale, J. (1993) *Immunol. Today* **14**, 349–352.
- 17 Trowsdale, J. (1993) *Trends Genet.* **9**, 117–122.
- 18 Jurka, J. and Milosavljevic, A. (1991) *J. Mol. Evol.* **32**, 105–121.
- 19 Iris, F.J.M., Bouguelert, L., Prieur, S., Caterina, D., Primas, G., Perrot, V., Jurka, J., Rodriguez-Tome, P., Claverie, J.M., Dausset, J., and Cohen, D. (1993) *Nature Genet.* **3**, 137–145.
- 20 Parimoo, S., Patanjali, S.R., Shukla, H., Chaplin, D.D., and Weissman, S.M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 9623–9627.
- 21 Lovett, M., Kere, J., and Hinton, L.M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 9628–9632.
- 22 Buckler, A., Chang, D.D., Graw, S.L., Brook, J.D., Haber, D.A., Sharp, P.A., and Housman, D.E. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 4005–4009.
- 23 Daniels, D.L., Plunkett, G. 3rd, Burland, V., and Blattner, F.R. (1992) *Science* **257**, 771–778.
- 24 Oliver, S.G. *et al.* (1992) *Nature* **357**, 38–46.
- 25 Wilson, R. *et al.* (1994) *Nature* **368**, 32–38.
- 26 Chen, E.Y., Schlessinger, D., and Kere, J. (1993) *Genomics* **17**, 651–656.
- 27 Smith, M., Holmsen, A.L., Wei, Y.H., Petersen, M., and Evans, G.A. (1994) *Nature Genet.* **7**, 40–47.
- 28 Grouthues, D., Voss, H., Stegemann, J., Wiemann, S., Sensen, C., Zimmermann, J., Schwager, C., Erfle, H., Rupp, T., and Ansorge, W. (1993) *Nucleic Acids Res.* **21**, 6042–6044.
- 29 Voss, H., Wiemann, Grouthues, D., Sensen, C., Zimmermann, Schwager, C., Stegemann, J., Erfle, H., Rupp, T., and Ansorge, W. (1993) *Biotechniques* **15**, 714–721.
- 30 Riley, J., Butler, R., Ogilvie, D., Finnear, R., Jenner, D., Powell, S., Anand, R., Smith, J.C., and Markham, A.F. (1990) *Nucleic Acids Res.* **18**, 2887–2890.
- 31 Isegawa, Y., Sheng, J., Sokawa, Y., Yamanishi, K., Nakagomi, O., and Ueda, S. (1992) *Mol. Cell Probes* **6**, 467–475.
- 32 Berg, C.M., and Berg, D.E. (1994) In *Mobile Genetic Elements: Frontiers in Molecular Biology*, Sherratt, D. (ed.). Oxford University Press, Oxford, UK, pp. 38–68.